**“XG-Boosting” of biofuel production in low N-by-W environments”**.

Gloria Coruzzi, Carroll and Milton Petrie Professor of Biology

New York University

(212) 998-3963, gloria.coruzzi@nyu.edu

Funding Opportunity Announcement Number: DE-FOA-0002214

**Relevance to Systems Biology and DOE-BER mission**: There is increasing interest in analyzing complex biological data using machine learning algorithms to infer a polygenic trait from genomic feature(s). We address these basic gaps in systems biology to address a practical application of increasing productivity in the model Biofuel crop Brachypodium in dry, nutrient poor soils. We use a suite of machine learning algorithms to predict Nitrogen Use Efficiency (NUE) and Water Use Efficiency (WUE) phenotypes based on transcriptome data in varieties of the model biofuel crop Brachypodium. The outcome will be a set of features (e.g. genes) whose importance to predicting phenotypes is ranked using out-of-sample crop varieties. The advantage of conducting these studies in Brachypodium, is that we can functionally test genes and uncover mechanisms involved in NUE and WUE by mutation and transgenesis. Our immediate application is to predict NUE and WUE in a model biofuel crop, but the approach we develop and deploy can be applied across biology, agriculture and medicine.

**Novelty: N-by-W gene responses predict WUE and NUE in crops.** Our recent studies in rice show that genes responding to N-by-W interactions (N/W or NxW), show the highest correlation with biomass and drought outcomes (e.g. W-use efficiency and Stomatal conductance) – compared to genes that respond only to W (Fig. 1) (2). However, to date, transcriptome studies in Brachypodium have been limited to drought responses alone (REFs). There are currently no datasets of the N-response transcriptome study in Brachy published to date. Our plan is to measure the N-by-W responses in Brachypodium and use them in a machine learning pipeline to uncover the genes that underlie NUE and WUE in the model biofuel crop, where we can functionally validate the mechanisms using mutants and transgenics.

**Overview of Aims: The goal of this study**, is to use machine learning approaches to uncover the genes that control biomass, WUE and NUE in the model biofuel crop, Brachypodium. In **Aim 1**, we will identify the genes and phenotypes responsive to N-moles, W-moles, N-molarity (N/W) and NxW synergistic interactions by exposing Brachy to a 4x4 NxW matrix, as we have previously done in rice (Swift 2019). Then, in **Aim 2**, we will extend this N-by-W analysis across a set of 15 Brachy varieties that have been identified to be tolerant, intermediate and susceptible to drought as described in (3). In **Aim 3**, we will use this N-by-W transcriptome and phenotype data in a collection of machine learning methods to which will identify a ranked list of features (e.g. genes) whose N-by-W responses are able to predict phenotype outcomes. Finally, in **Aim 4**, we will functionally validate the role of these genes in the N-by-W responses using mutants and transgenics in phenotypic assays.

**Background**:

**N-by-W interactions control plant growth, development and Biomass.**

Since both N and W are increasingly limited in soils world-wide, N-based fertilizers and irrigation underlie modern agriculture’s goal to meet yield potential. In the coming decades, climate change will force farmers around the globe to adapt to drier, nutrient-poor soils(4, 5). At the same time, the damaging environmental impacts of synthesizing N-fertilizers - including disruption of the global N-cycle - are already being felt (5). Consequently, in an attempt to develop crops that are either N- or W-use efficient, research efforts have focused on understanding how the availability of N *or* W *individually* in soils impacts plant biology - both at the physiological and molecular level (2, 6-9).

Also, in recent years, the molecular and signaling components that underlie plant responses to either N or W have begun to be elucidated (10, 11); however how they overlap with one another remains poorly understood. Since combinations of N and W have a clear impact on important plant traits, it is likely that there are synergistic effects between the molecular mechanisms that sense and respond to N and W. **To address this knowledge gap**, we propose research strategies to study how genes responding to N-by-W interactions contribute to improving plant growth on arid, nutrient-poor marginal soils.

**RESULTS**

**N-by-W interactions are crucial to both WUE and NUE in RICE.** Given that water (*W)* is a solvent for nitrogen (*N)*, plant responses to these two key nutrient resources must be tightly linked. The study by Swift et al., 20191 was the first to design and test a matrix of *N* and *W* doses that could test whether and how plants sense *N*-dose independent of water (e.g. moles), or the ratio of *N*/*W* (e.g. molarity) (Fig. 1A). Using expression data from RNA-seq, linear models were used to uncouple the genes that responded specifically to four different *N*-by-*W* categories, *N* (moles), *W* (volume), *N/W* (molarity) and *N*x*W* (molar synergy) in the lab and field1. For each category of *N*-by-*W* responses, the genes that showed consistent expression patterns (up or down) in both lab and field studies were used to derive eigengene values. Eigengene expression values that correspond to *N*-by-*W* interactions (*N*/*W* or *N*x*W*), correlate with beneficial rice phenotypes such as grain yield and water use efficiency (Fig. 1B).





***RESEARCH PLAN***

***Aim 1.* Identify N-by-W response genes in Brachypodium. *Rationale*:** Our studies of N-*by*-W sensing mechanisms in rice have uncovered novel principles of nutrient-dose sensing in biology that affect crop biomass and yield in the field (2). Specifically, we discovered that rice plants can sense and distinguish between N-moles, W-moles, N-molarity (N/W), and the synergistic interaction of ‘NxW’ - an effect we dub ‘molar synergy’ (**Fig. 1A**). Through field tests of 19 rice varieties, we found that gene sets responding to N and W combinations - molarity (N/W) and molar synergy (NxW) - could predict agronomic crop traits (2) (**Fig. 1B**). **Our goal in Aim 1** is to apply this approach to acquire new basic knowledge of how N and W interactions impact gene expression and phenotypic outcomes in the model *biofuel species, Brachypodium*.

**Aim 2. Test N-by-W response phenotypes and transcriptome across 15 Brachypodium varieties with varying drought sensitivity.** This aim will test the N-by-W responses – phenotypes and transcriptome- using a 2x2 NxW matrix, as we did for rice varieties (2). We will perform this using a select panel of 15 Brachy lines which were classified as tolerant (TOL), intermediate (INT), or susceptible (SUS) to the drought conditions imposed during the drought screens (3). The data generated in this aim, will be used to fuel the machine learning pipeline in Aim 3 - which will rank genes based on their importance in being able to predict the phenotypes -e.g. biomass, NUE, and WUE of Brachy lines based solely on gene expression.

**Aim 3. Using ensemble machine learning methods to predict phenotypes based on N-by-W responsive genes.** We present a pipeline to infer a complex trait from transcript abundanceusing the matched phenotypic and transcriptomic data generated in Aim 2. We recently successfully used the conserved N-response data from Arabidopsis and maize to infer NUE from transcript abundance using a EXtreme Gradient Boosting algorithm (XGBoost). In this aim we will expand and optimize the machine learning pipeline using a suite of linear and non-linear methods because no one single algorithm always works the best for all species and traits (12). We will partition the datasets into training (n-1 genotypes) and testing (the left-out genotype) (Fig. 2A) and use the training data to optimize the NUE models (Fig. 2B). Next, we will use the trained models to predict NUE for the left-out genotype. To evaluate the model performance, we will compute the correlation between actual NUE in the test set and the predicted NUE. To prioritize the validation in Aim 4, we will rank the ‘important features’ (i.e. genes) from the top performing algorithms. This pipeline will also be run independently for WUE traits. 

**Aim 4: Validation of candidate genes for adaptation to growth in low-N/low W soils.** The result in Aim 3 will identify a set of ranked important features (genes) whose expression values predicts traits such as biomass, NUE, and WUE. The top-ranked genes from this list will be functionally validated in Brachypodium at the Wisconsin facility XXXX. The functional validation assay will generate transgenic lines that constitutively over-express the candidate gene or mutant lines lacking it, we will grow each line in soil with low-N content spiked with 15N to measure NUE. Candidate genes that produce greatest increase in NUE and biomass relative to the wild-type, will be top candidates for introduction into biofuel crops to increase their biomass accumulation in marginal N-poor soils.

**Outcome:** This team of scientists who have an established history of extensive collaboration will train students & post-docs in the fields of genomics and machine learning. This project will also provide genes that when introduced into biofuel crops will greatly improve their WUE and NUE in arid, N-poor marginal soils.

1. Xu C, Jackson SA. Machine learning and complex biological data. Genome Biology. 2019;20(1):76.

2. Swift J, Adame M, Tranchina D, Henry A, Coruzzi GM. Water impacts nutrient dose responses genome-wide to affect crop production. Nature communications. 2019;10(1):1374.

3. Fisher LH, Han J, Corke FM, Akinyemi A, Didion T, Nielsen KK, et al. Linking Dynamic Phenotyping with Metabolite Analysis to Study Natural Variation in Drought Responses of. Front Plant Sci. 2016;7:1751.

4. Ulrich W, Soliveres S, Maestre FT, Gotelli NJ, Quero JL, Delgado‐Baquerizo M, et al. Climate and soil attributes determine plant species turnover in global drylands. Journal of biogeography. 2014;41(12):2307-19.

5. Robertson GP, Vitousek PM. Nitrogen in agriculture: balancing the cost of an essential resource. Annual review of environment and resources. 2009;34:97-125.

6. Di Paolo E, Rinaldi M. Yield response of corn to irrigation and nitrogen fertilization in a Mediterranean environment. Field Crops Research. 2008;105(3):202-10.

7. Shi J, Yasuor H, Yermiyahu U, Zuo Q, Ben-Gal A. Dynamic responses of wheat to drought and nitrogen stresses during re-watering cycles. Agricultural Water Management. 2014;146:163-72.

8. Brueck H, Senbayram M. Low nitrogen supply decreases water‐use efficiency of oriental tobacco. Journal of Plant Nutrition and Soil Science. 2009;172(2):216-23.

9. Patterson T, Guy R, Dang Q. Whole-plant nitrogen-and water-relations traits, and their associated trade-offs, in adjacent muskeg and upland boreal spruce species. Oecologia. 1997;110(2):160-8.

10. O'Brien JA, Vega A, Bouguyon E, Krouk G, Gojon A, Coruzzi G, et al. Nitrate transport, sensing, and responses in plants. Molecular plant. 2016;9(6):837-56.

11. Vishwakarma K, Upadhyay N, Kumar N, Yadav G, Singh J, Mishra RK, et al. Abscisic acid signaling and abiotic stress tolerance in plants: a review on current knowledge and future prospects. Frontiers in plant science. 2017;8:161.

12. Azodi CB, Bolger E, McCarren A, Roantree M, de Los Campos G, Shiu SH. Benchmarking Parametric and Machine Learning Models for Genomic Prediction of Complex Traits. G3 (Bethesda). 2019;9(11):3691-702.